Insights from self-organizing maps for characterizing accessibility to healthcare networks

Jerome R. Mayaud[1]*, Sam Anderson[2], Martino Tran[1], Valentina Radic[2]

[1]School of Community and Regional Planning, University of British Columbia, 6333 Memorial Road, Vancouver , Canada **jerome.mayaud@gmail.com*
[2]Department of Earth, Ocean and Atmospheric Sciences, 2020–2207 Main Mall, Vancouver, Canada

1 Introduction

As urban populations continue to grow worldwide, it becomes increasingly important to evaluate network accessibility (the ease with which residents can reach key places or opportunities). In an effort to become more liveable and sustainable, 'smart cities' are seeking to optimize accessibility to essential services such as healthcare, while ensuring equity is maintained amongst different population groups. For the first time, we apply a powerful machine learning (ML) tool, the self-organizing map (SOM), to cluster income characteristics and relate them to accessibility to healthcare service networks across the rapidly growing, 'smart' City of Surrey, Canada. We perform this analysis for both 2016 and 2022, to examine the potential shifts in accessibility over time.

2 Methods

We use a SOM to reduce the dimensionality of our income datasets for 2016 and 2022 and draw out patterns in the underlying data to relate them to accessibility metrics. First, we used open-source data to evaluate accessibility across Surrey. We divided the surface area of the city into a grid of 1,480 equally sized hexagons (500 m diagonal diameter), following the method of [1]. We then calculated travel-time estimates (optimally combining public transportation and walking) between every pair of grid cells – an 'origin' and a 'destination' (O-D) – using OpenTripPlanner (OTP). The spatial layout of road networks and pedestrian infrastructure used in OTP was acquired from OpenStreetMap, and public transport routing was derived from geolocated timetables for a typical commuting day in September 2017.

We combined our O-D matrices with high-resolution (Dissemination Area) census population data from 2016 and population projections for 2022 [2]. The census data were spatially reorganized into the aforementioned hexagonal grid. Since census income data are reported in bins, we used MATLAB's built-in '*ksdensity*' method to perform a kernel density estimation (KDE) that provided us with a smoothed, normalised continuous probability distribution for income in each grid cell. These distributions were used as the input for our SOM algorithm. The population counts for each grid cell were used to perform catchment area analysis using the O-D matrices. Accessibility for each origin grid cell (for a total n grid cells) was calculated as $F_{o,T} = \sum_{d=1}^{n} F_d \, f(t_{odr})$, where $F_{o,T}$ is the number of facilities F that can be reached from origin o within time threshold T (30 mins in this study), F_d is the number of facilities in destination cell d, and $f(t_{odr})$ is a time threshold function whose value (0 or 1) depends on whether t_{odr} is greater or smaller than T. We also calculated the travel-time to the nearest facility for each grid cell.

The essential concept behind SOM analysis is to cluster observations onto a two-dimensional topology of nodes (patterns) presented in a regular 'map' [3]. We created a SOM through an iterative 'training' process that compares input data samples to each SOM pattern, and computes the minimum Euclidean distance to determine the closest match for each sample [4]. We quantified the magnitude of income distribution change between 2016 and 2022 by performing principal component analysis (PCA) to determine the first two modes of each SOM pattern distribution in both study years. A 'cluster distance' metric was then derived from the Euclidean distance in PC-space between the 2016 and 2022 income cluster patterns.

3 Results

Our method yielded an 8-cluster (4 rows x 2 columns) SOM (Fig. 1). We found that income homogenization is projected to occur over the study period: higher-income clusters (e.g. cluster 1 in red) will become more prevalent by 'absorbing' members from other clusters. This will result in decreased average accessibility for the absorbing clusters, and vice versa. By 2022, SOM cluster transitions and demographic shifts will result in ~12,000 more seniors residing in areas with no access to a hospital and ~1,500 with no access to a walk-in clinic.

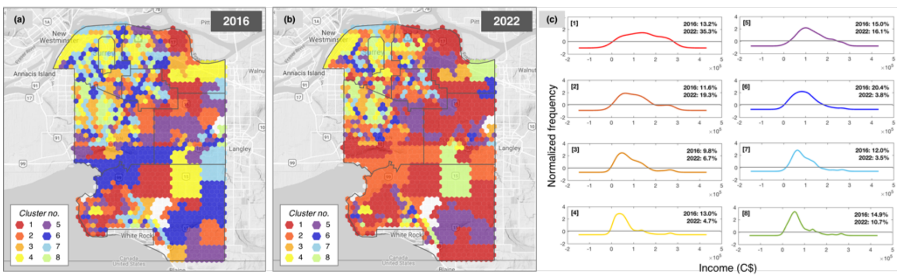

Fig. 1: Spatially mapped SOM topology, coloured according to clustering, for (a) 2016, and (b) 2022; (c) Frequency distributions for each SOM cluster, showing in bold the frequency that they occur in the 2016 and 2022 maps.

Cluster distance analysis for both hospitals and walk-in clinics (Fig. 2) suggests that ageing populations (i.e. redder colours) will be concentrated in areas with relatively poor hospital and clinic accessibility, although income distributions in these areas will remain relatively stable. In contrast, median age will decrease (bluer colours) in areas experiencing large changes in income distribution and/or with access to numerous healthcare facilities.

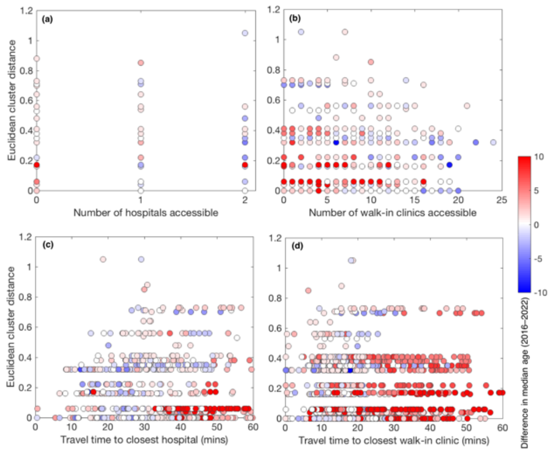

Fig. 2: Scatterplots showing relationships between Euclidean cluster distance and access to healthcare facilities (in terms of the number of facilities and travel-time to the nearest facility) for each cell. A third coloured variable shows the difference in median age between 2016 and 2022.

4 Conclusion

Our results suggest that a dual accessibility problem may soon arise in Surrey. First, large senior populations will reside in areas with access to numerous, and close-by, clinics, which will put pressure on existing facilities for specialized services. Second, lower-income seniors will increasingly reside in areas poorly connected to healthcare service networks; these populations are likely to be reliant on public transportation, so accessibility equity may suffer. As fertility rates decline and life expectancies increase around the world, shifts to elderly, less mobile populations will pose major challenges for healthcare accessibility in ageing cities like Surrey.

Our study confirms there is much potential in applying SOM and other ML methods more widely in fields related to accessibility and urban networks. SOMs in particular help to uncover and visualise data patterns in ways that are communicable and relevant to policymakers.

References

[1] Mayaud, J.R., Tran, M., Pereira, R.H.M., Nuttall, R. Future access to essential services in a growing smart city. *Computers, Environment and Urban Systems,* (2018)

[2] Statistics Canada. http://www12.statcan.gc.ca/ (accessed November 2017)

[3] Kohonen, T.: Self-organizing Maps. Springer, Berlin, 3rd Edition (2001)

[4] Unglert, K., Radic, V., Jellinek, A. M. Principal component analysis vs. self-organizing maps combined with hierarchical clustering for pattern recognition in volcano seismic spectra. Journal of Volcanology and Geothermal Research, 320, 58–74 (2016)